



Outlook for Navigation - Comparing Human Performance with a Robotic Solution

Blanke, Mogens; Hansen, Søren; Stets, Jonathan Dyssel; Koester, T. ; Brøsted, J. E. ; Maurin, Adrian Llopert; Nykvist, Nicolai; Bang, J.

Published in:
ICMASS 2018. Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Blanke, M., Hansen, S., Stets, J. D., Koester, T., Brøsted, J. E., Maurin, A. L., Nykvist, N., & Bang, J. (2019). Outlook for Navigation - Comparing Human Performance with a Robotic Solution. In *ICMASS 2018. Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships* SINTEF.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Outlook for Navigation - Comparing Human Performance with a Robotic Solution

M. Blanke¹, S. Hansen¹, J. D. Stets², T. Koester³, J. E. Brøsted³, A. Llopart Maurin¹, N. Nykvist¹, J. Bang⁴

¹ Technical Univ. of Denmark, Dept. of Electrical Engineering, {mb,sh,addlo}@elektro.dtu.dk

² Technical Univ. of Denmark, Dept. of Applied Mathematics and Computer Science, stet@dtu.dk

³ Force Technology, {tsk,jeb}@force.dk

⁴ Danish Maritime Authority, cjb@dma.dk

Corresponding author e-mail: mb@elektro.dtu.dk

Abstract

Considering whether a temporarily unattended bridge could be allowed, Maritime Authorities wish to investigate whether sensor technology is available that, when seconded by sophisticated computer algorithms, is able to provide outlook with the same reliability and safety as that of the average human outlook. This paper reports findings from a comparative study of human versus electronic outlook. Assessment of navigator's outlook is based on measurements with a wearable eye-tracker and areas of their visual attention are recorded on video. Simultaneously, a set of electro-optical sensors provides image-data as input to computer algorithms that detect and classify objects at sea within visual range. The paper presents the methodology used to deduct, from the observations of fixations, when the navigator turned his attention to a particular object and compares this with the Electronic Outlook. On the technology side, the paper details on how machine learning is used for object detection and classification, and discusses quality attributes, including efficiency and robustness of detection and classification, expressed through statistical measures.

Keywords: Outlook for navigation, autonomous vessels, electronic outlook, human outlook.

1. Introduction

Look-out for navigation is the task of observing various objects which can have an impact on a ship's planned route and maneuvering capabilities, for example other vessels, buoys and land. If the outlook is a separate person on the bridge, observations are reported to the officer in charge who decide any remedial actions. The look-out is made using sight and aided by available technology such as RADAR, AIS and ECDIS systems. Development within camera technology and computer vision algorithms has provided an additional possible source for look-out. This study investigates the quality of this "electronic outlook" and compares with human look-out.

A survey of maritime object detection and tracking methods was published in the survey by [21], who emphasized that RADAR, which is required by IMO on merchant vessels, is sensitive to the meteorological condition and the shape, size, and material of the targets. They emphasize that RADAR data need to be supplemented by other situational awareness sensors to obtain safe navigation and collision avoidance. Electro-optical sensors were available in this study for several spectral ranges: visual (450-800 nm), near infrared, (NIR 800-950 nm) and long wave infrared (LWIR 8-14 μm). Outlook was based on eye-tracking by glasses that monitor the Navigator's areas of attention, judged by observed *fixations*. The eye-tracking glasses were limited to determine fixations on outside bridge objects in daylight conditions, and this defined the scope of comparison in this paper.

The paper first summarizes the task of watch keeping/lookout for navigation in Section 2, and 3 explains how human outlook is observed through measurements where a navigator wears eye-tracking glasses. Section 4 outlines the use of electro-optical and other sensors to provide

electronic means to replicate the human observation of surroundings. Section 5 introduces present technology for object detection and classification at sea, showing the features obtainable with image processing and machine learning techniques, while Section 6 provides details on data and training. Section 7 presents results on object detection performance for the network chosen. Section 8 presents findings from ferries in near-coastal and shallow water navigation and Section 9 discusses limitations and perspectives of results. Finally, conclusions and future directions are offered in 10.

2. Outlook for navigation

A. Human outlook

The analysis of manual lookout/watch-keeping is based on a combination of observations on board several vessels in Danish waters. Electronic observations and Eye tracking measurements were conducted during the summer of 2018 on ferries in Northern Øresund and South Funen archipelago.

Further, but outside the scope of this study, generic observations were made on board a large number of vessels during the period 2000-2018. The generic experience also includes observations from ship simulator exercises at FORCE Technology in Lyngby, general knowledge on human factors as well as literature, see [25] and [27].

B. Endogenous and exogenous driven visual attention

The look-out task involves both endogenous- and exogenous-driven activities. Endogenous activities are visual attention controlled by the navigator himself on his own initiative and based on relevant knowledge and experience, such as observing navigational markings, sighting

of land and watching out for other vessels. Exogenous activities are caused by an external (and in principle unforeseeable) event catching the attention of the navigator. For instance, the sight of a vessel which the navigator has not been looking for or some light or sound signals. Everyday scenarios will typically be a combination of endogenous and exogenous look-out activities.

It is important to be aware that the outlook is just one among several tasks of the navigator on the bridge. Other tasks include observation of the condition of engines and systems, communication and passenger and safety related tasks.

When it comes to performing an outlook, it makes sense to distinguish between pure observations not requiring action and observations requiring action, e.g. to prevent a collision. An action is often seen as a combination of several elements including signalling, steering and engine manoeuvres, but the decision to act could not be covered by the present analysis.

1) *Recognition of objects*: The navigator's recognition of objects is based on both the visual appearance and on the behaviour of objects.

This study has not employed means to disclose how the navigator interprets what he sees. The eye tracking glasses can determine where the navigator has had visual focus. The detailed recognition of objects and their behaviour are therefore not in the scope of this investigation.

3. Eye-tracking

In the maritime context, the use of eye tracking as means to examine the visual attention of ship navigators is nothing new. At least not when it comes to the use of eye tracking in simulation environments. [3] investigated the operators' foci of attention during simulated dynamic position operation. [2] examined the difference in attention-allocation comparing novice and expert navigators during use of the Conning Officer Virtual Environment, a simulation system developed to train ship handling. [2] concluded a clear link between the experts' superior ship-handling performance and a "tight Attention-allocation pattern that focused only on the relevant areas of interest. Novices' Attention-allocation patterns were highly scattered and irregular" (p. xviii). [19] and [23] focused on evaluating and improving the training of navigators using eye tracking data and [20] suggested using (stationary) eye tracking to determine or monitor the level of fatigue in the boat driver with the purpose of enhancing situation awareness. [11] used eye tracking data examination to suggest improvement of usability design on the ships' bridge layout and in the software's graphical user interface on a maritime navigation display. [12] also investigated eye tracking data in the pursuit of a recommendable optimal *visual scan pattern* for navigators aiming to mitigate the mental workload needed to monitor the increasing amount of technology used at ship's bridge.

A somewhat rare example of an investigation using eye tracking during actual, real life navigation was presented in [8]. They investigated gaze behavior data from 16 experienced and novice boat drivers during high speed navigation and concluded that novices looked more at

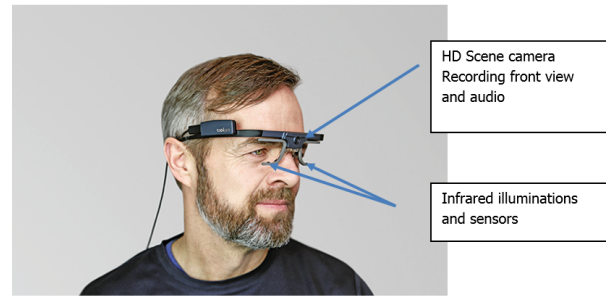


Fig. 1: Tobii® eye tracking glasses. (photograph courtesy of FORCE Technology)

objects closer to the boat while experts looked more at things far from the boat. Also, novice boat drivers were more focused on electronic displays, while the experts focused mostly outside the boat and used the paper-based sea chart to a larger extent than novice drivers.

The methodology of using eye tracking devices in real life maritime situations is not often seen, and is considered a feature of this study.

A. Eye tracking technology applied in this investigation

The eye tracking data was collected using Tobii® Pro Glasses 2 ([1]), which is a lightweight wearable technology illustrated in Figure 1

The head unit has a scene camera recording the wearer's front view (including audio) and the frame has infrared illuminators and sensors installed thereby using the eye tracking technique Corneal reflection (dark pupil). The belt clip unit holds a SD card for recording data, operates on rechargeable batteries and is Wi-Fi controlled through PC-based software (in this case iMotions®). This setup makes it very easy for the person wearing the eye trackers to freely move around on the ship and due to the non-invasive design, most subjects easily forget they are even wearing them while performing their job. Additional specifications are shown in the table below, adapted from the Tobii Pro Glasses 2 User's Manual (2018, p. 40). Based on the recording from the scene camera and the associated eye tracking data, the iMotions software (version 7.1) produces a video showing what was in the wearer's field of view during the recording (a 1st person perspective replay), including a graphical overlay. A yellow dot indicates where the person was looking at any given time, within the field of view. The software was set to illustrate fixations by increasing the size of the yellow dot. A fixation is defined as a period (100 ms or more) in which the person's eyes are focused on a specific object (or location) in the field of view. Fixations are excellent measures of visual attention [14], [19].

The image in Figure 2, shows a single frame from replay of an eye tracking recording. The yellow dot is the location of the navigator's fixation and the yellow line illustrates eye movements faster than 100 ms (ie. saccades).

B. Limitation in scope due to equipment

The eye-tracking technology was challenged by the high contrast between outdoor and inside bridge, and eye-



Fig. 2: Eye tracking example in dense traffic and confined from South Funen archipelago.

tracking could not reveal which objects on the Radar screen or on the ECDIS caught the attention of the navigator. Eye tracking could not be used in low-light conditions during dusk and evening. The electronic to human outlook investigation was therefore restricted to compare performance in daylight conditions.

4. Electronic outlook

The electronic outlook system in this comparison consist of 5 cameras, an FMCW RADAR and an AIS receiver for reference. The vision system is composed of 2 colour cameras (JAI GO-5000C 2560×2048 , 12 bit), 2 monochrome cameras (JAI GO-5000M, 2560×2048 , 12 bit) with longpass filters for the NIR range and 1 LWIR camera (Teledyne Dalsa Calibir 640, 640×480 , 14 bit). The sensors are mounted on a forward facing stand on board, see Figures 3 and 4.

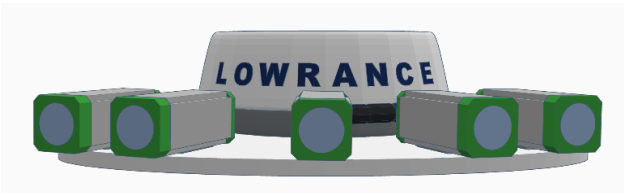


Fig. 3: Sketch of the sensor platform. The five camera houses are looking forward. Camera units, CW-FM RADAR and GPS receiver are mounted on the test platform. The combined horizontal field of view of two daylight cameras is 110 deg.

5. Object detection and classification

We wish to identify what objects are present on the water within a given distance from our vessel. Information about stationary objects such as buoys, rocks, bridge pillars and islands, and moving objects such as boats, ferries, etc. are important for positioning, navigation and collision avoidance.

A. Image-based Object Detection

We use image-based object detection and classification to determine what is present in the environment in which we navigate. Our electronic outlook system is continuously



Fig. 4: Southern Funen archipelago. Sensor platform mounted beyond wheelhouse / ship's bridge.

sampling images at a fixed rate, and we wish to know what objects are present in the images and where. This is valuable information that can later be used to determine the objects approximate position relative to our vessel.

For this task we use instance segmentation, which is a pixel-wise classification of the image. Using instance segmentation, we not only get classifications of the objects present but a segmentation mask of each of the instances in the image i.e. if more objects of the same class are present in the image, each of them are assigned a unique label. That enables us to potentially track individual objects from the same class.

Recently, data-driven solutions, such as deep neural networks, have proved to give robust and accurate results but these require large sets of annotated training data. Annotations often have to be done manually, and especially pixel-wise annotations for semantic and instance segmentation requires accurate annotations which can be cumbersome. Techniques that require less or no prior data also exist but tend to be less generalizable than a learning-based approach. Since our system is operating near the coast, many types and sizes of boats and ships can appear in the images. Additionally, we can have both land and water as background. The following provides an outline of some challenges for a maritime environment along with related prior work.

B. Related work

Several previous works address object detection, classification and tracking in a maritime environment. Challenges include waves that can cause a rapid change in the frame of reference [7], sudden change of illumination and unwanted reflections from the water [4], and the possibility of poor weather conditions that reduce the range of sight. As mentioned in the survey papers [21], [18] there exist a range of methods concerning detection and classification in images of the maritime environment, and horizon line detection and background subtraction seems to be effective for object detection [28], [26]. Methods include to utilize infrared and visible light images [21], but also thermal imaging alone has the ability to provide information about objects on the water [16]. With recent progress in deep

learning based segmentation and classification methods, visible light images is an obvious choice for object detection since much training data, such as e.g. ImageNet [6], already exists and can provide a good base for training. Specifically for maritime environments, [15] and [5] show that deep learning methods are effective, and annotated data from the maritime environment exists [21]. This project has used training data collected from observations on-board ferries in Danish coastal waters.

C. Mask-RCNN detection and classification

Objects that are within visual range of the cameras are detected and classified using a Convolutional Neural Network (CNN), also referred to as deep learning technology. The network architecture employed in this project to detect different objects in the maritime environment is Mask-RCNN [13], which has the novelty of not only being able to recognize and detect (bounding box) of several classes, but is also able to segment all instances of each one and create the corresponding binary mask at a pixel level. Mask-RCNN is an architectural model that started with a Region-Based Convolutional Neural Network (RCNN) [10], followed by Fast-RCNN [9] and then Faster-RCNN [22].

6. Dataset and Training

We found that existing maritime image datasets are not sufficient to cover the scenarios we encounter in our recordings. Consequently, a subset of images is hand-annotated and used for both network refinement and to test the performance of the detection algorithm. The subset is labelled for instance segmentation so that pixels belonging to each object in the image is labelled separately with a polygon shape. Manually labelling of images for instance segmentation is a time consuming and to ease the process we use a free web-based annotation tool *LabelMe* [24] to create polygons. Each object is assigned to a class and Figure 5 shows how polygons are drawn for each object in a picture. The process of manual labelling an image with a few objects takes from 1-5 minutes depending on the complexity of the silhouettes.

The images annotated were captured with the on board RGB camera setup and additional images were acquired with a DSLR camera on separate trips. Images from internet sources are also added to the training data. All images were manually annotated using the above mentioned technique. In summary, the annotated images for the data-set consists of:

Data source	Number of images
On-board RGB camera setup	330
On-board DSLR	179
Internet source	8
In total	517

The 517 images are annotated with two classes: buoy and ship. A total of 600 buoys and 639 ship instances are annotated across the data-set.



Fig. 5: Green polygons show the boundaries for one boat and two buoys that are present in this image.

A. Training

The on-board RGB images are split so that 406 images are used for training and 111 are used for validation. The validation set consists of images from the on-board RGB camera setup, as we wish to evaluate the performance of the object detection on the on-board camera system. To produce additional training data, data augmentation was used on each of the on-board RGB training images as follows: random rotation within a $\pm 25^\circ$ range, flip image horizontally (mirroring), combine flipping and rotation, replace an image pixel with a chosen colour for every 50 pixels.

The augmentation increases the data-set with an additional 5×406 images. The images are cropped into 16 regions in a 4×4 grid. After this operation, the total increase of the data-set is $16 \times 5 \times 406$ images, resulting in $16 \times 5 \times 406 + 406 \times 5 = 34510$ images.

The Mask-RCNN uses the pre-trained weights obtained from the COCO dataset [17] and we fine-tune the network to detect the two classes provided in our training data: buoy and ship. The network was trained for 40 epochs on the first 4 layers (classificatory), then another 60 epochs for the rest of the layers and finally 80 epochs for the whole network. The learning rate was set to 0.0003 and the momentum was 0.9. The total training time took around 24 hours on a GeForce GTX 1080 GPU.

7. Performance

This section evaluates the performance of the network used through validation of images from the on-board RGB camera system. With the above-mentioned training procedure, we obtain a mean average precision (mAP) of 62.74%. The 0.5-mAP is used which means that intersections of regions less than 50% are not included in the calculation.

Object detection is done in two stages. First, detect and classify a relevant object in the image. Second, determine how accurately it is segmented. To discuss the results with the aim of supporting navigation, the mean average precision (mAP) is not very useful as a measure of quality. The reason is that safe navigation requires that

Reference		Detected & Classified			
		Buoy	Ship	~ Buoy	~ Ship
near	Buoy	47	0	0	
	Ship	0	83		0
far	Buoy	27	1	54	
	Ship	0	51		0
none	~ Buoy	6			
	~ Ship		34		

TABLE I: Performance of the object classification. Detected objects are compared to objects that were labelled in the validation set. The number of detections is noted for two categories of objects: buoy and ship. The distance to objects are divided into near and far. The symbol \sim denotes negation.

all objects are detected, which might present a risk to safe navigation. We therefore employ the standard terminology from statistics for quality assessment of object detection and classification:

True positive Object is present in a frame and is detected.

False positive Object is not present but a detection occurs.

True negative Object not present in the frame and no detection occurs.

False negative Object present in the frame but is not detected.

For our application, we need a good overall localization of the object in the image, but not necessarily a precise segmentation border around the object. We conclude that segmentation of the objects are acceptable in most cases where a true positive detection occurs, using visual inspection.

We also wish to investigate to what extent the network is detecting the objects it is supposed to find, the occurrence of false positives i.e. false classifications. To do this we note down the comparison of the reference (ground truth) annotations with the predictions provided by the network. The precision of the segmentation mask is omitted here, so it is only the object classification which is reflected in this part of the results. Note that our validation set consists of annotated images with one or more objects, but also images without objects are included in the set. Table I shows the results of the object detections and classifications. We consider the two object classes buoy and ship and divide the detections as near and far. The separation near versus far was determined by the estimated distance to an object in the frames.

The results in Table I show data for the validation set. Classification of nearby objects is very satisfactory. 100% of buoys and 100% of ships are found, and none are misclassified. With objects farther away numbers drop to 33% correct classification of buoys and 66% of ships. One buoy is detected but is misclassified as a ship. No ships are mistaken for buoys. False positives occur at far distance, a total of 6 buoys and 34 ships were detected without being present.

The numbers in Table I are valid for single frame recordings in the validation data set. Since the relative distance to objects are reduced as they approach, they

are eventually detected and classified. The essence is that objects are detected and classified in time to plan a path for safe navigation and collision avoidance. Whether detection and classification of far away objects is critical therefore depends on time to encounter.

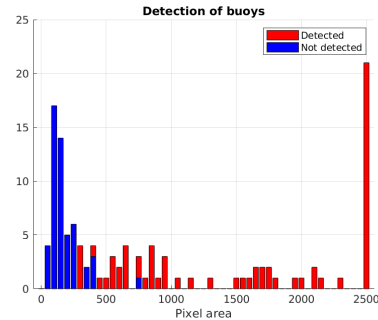


Fig. 6: Histogram of pixel area versus buoy detections.

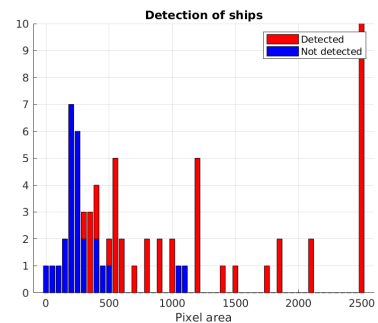


Fig. 7: Histogram of pixel area versus ship detections.

The false positives are often detections on the water where a piece of land far away is detected as a ship or in the region above the horizon line, where clouds are detected as ships. While it is not entirely straightforward task, we argue that a number of false positives in the cloud region could be removed by detecting the horizon in the image, as part of a robustification of the classification.

Classification performance is further scrutinized in Figures 6 and 7, which show missed detections in blue and correct detections in red colour as a function of pixel area. The Figure reveals that probability of detection raises sharply when object size in the image is above 450 pixels. All objects larger than 2500 pixels are detected but are not shown in these histograms.

Object classes were limited to *buoy* and *ship* to take advantage of the more than 3000 images of the class *ships* from the COCO data-set. For assessment of properties of the objects met at sea, it would be an advantage to add more classes to cover navigation and manoeuvring capabilities of nearby objects.

It is noted that the above statistics are based on inspection in the visual range only. Additional sensors such as near infrared and thermal imaging provide additional valuable information, but have not yet been included in the classification pipeline in this stage of the study.

8. Results

This study compares the human outlook by assessing the fixations determined by the eye-tracking system with object classifications made by the electronic outlook. Eye-tracking glasses were unable to determine areas in focus on RADAR or on the electronic chart display (ECDIS) screen on the bridge.

Comparison between the capabilities of electronic outlook and the human counterpart are therefore done looking at the instant of first observations of a given object. The eye-tracking software gives an indication of fixation on an object when the human lookout has been gazing at it for 100ms. This time is compared to the time-stamp that the Mask-RCNN indicates its first detection and classification of the object. Figure 8 shows a snapshot of eye-tracking. The right part shows what the lookout is focusing on. The yellow line on this shows that the eye focus wander around, which is normal. Fixation is indicated by the red circle. The Electronic Outlook is illustrated in Figure 9.

A. Temporal Comparison

This section presents an analysis of the time-wise differences between the electronic lookout system and the human counterpart. This is achieved by time-stamping detection of objects observed by the electronic lookout and comparing them with fixations captured by the eye-tracking system. A comparison is done by examining the difference

$$\Delta t_{obs} = t_{HO} - t_{EO} \quad (1)$$

where t_{HO} is the time that the eye-tracking system indicates the first fixation on an object, and t_{EO} is the time that the electronic outlook first detects and classifies the same object. Figure 10 shows a histogram of Δt_{obs} . Figure 11 shows the time difference Δt_{obs} histogram for ships and buoys separately. A positive value of time difference means that electronic outlook classifies an object earlier than the navigator has a fixation on it.

The time elapsed between the instant of detection of an object and the instant when this object passes behind the RGB camera's field of view is defined as the time to react. Two time differences are defined to analyze this characteristic,

$$\Delta t_{HO} = t_{pass} - t_{HO} \quad (2)$$

$$\Delta t_{EO} = t_{pass} - t_{EO} \quad (3)$$

where t_{pass} is defined as the time instant when the object passes behind the RGB cameras' field of view.

Figure 12 shows Δt_{HO} vs Δt_{EO} . The range is 0–200 s before passing own vessel. In average, electronic outlook allows more time to react.

9. Discussion

Since the ship has a RADAR and AIS sensors on board, the detection of objects that are visible to RADAR or have AIS transmitters, could be done quite accurately. However, several objects are not visible on RADAR, such as leisure surf borders and sea kayaks, boats without

RADAR reflector and AIS transmitter, and even containers that accidentally dropped over board. Electronic outlook with object classification is therefore essential for the ship to act in a safe manner.

Object detection performance of the Mask-RCNN network showed a satisfactory detection probability for objects larger than 400-500 pixels in an image, a quantification that is useful for camera system design for electronic outlook. However, a few outliers exist in the form of some false detections and very few missed detections. Missed detections can be critical and are believed to be a consequence of lack of training of the network. Sufficient coverage in the training of a neural network, and robustness of detection, are challenges that need be further addressed. A combination of neural net classification with more classical image analysis methods, addition of object tracking, and fusion with other sensor information could be ways to obtain robust classification.

A combination of object positions from these sensors and the Mask-RCNN architecture could increase the performance and the results. Examples include object tracking from camera information and using detected objects positions, by vision sensors and by Radar, as possible region proposals in the network.

Further results will, therefore, fuse on-board RADAR and AIS information with visual information in different spectral ranges. This will include calibration that enables RADAR and AIS data to be projected into e.g. the pixel-coordinates of the input images to the CNN. This data could be used for region proposal in the network and be particularly useful in situations with reduced visibility of the cameras.

A. Coverage of this analysis

Some of the elements of look-out are not captured by only observing the fixtures with eye tracking glasses, but would require further interpretation. This includes: general visual observation of nothing in particular, but often focused on the direction of the vessel and abeam/passed objects in relation to progression of the navigation; exogenous-oriented attention – something turns up - can include comparison or verification with information from Radar and AIS; endogenous-driven observation of objects from other sources – sea charts, Radar or AIS .

Such interpretation of the situation was not part of this study.

B. Electronic outlook as a fifth sense supplement for the navigator

Look-out is just one among several tasks of the navigator on the bridge. Other tasks include: observation of the condition of engines and systems; handling of cargo and passengers; safety-related routines; communication internally on board the vessel and with external parties; management of staff and other administrative tasks; QA and documentation tasks; handling of safety-critical situations on board.

With several other tasks to care for, which might sometimes distract the navigator, it is believed that electronic



Fig. 8: Eye-tracking of the manual look-outs fixations. Left: Forward facing camera used as reference in the analysis. Right: Eye-tracking result. The yellow spot surrounded by a thin red line indicates fixation on an object.

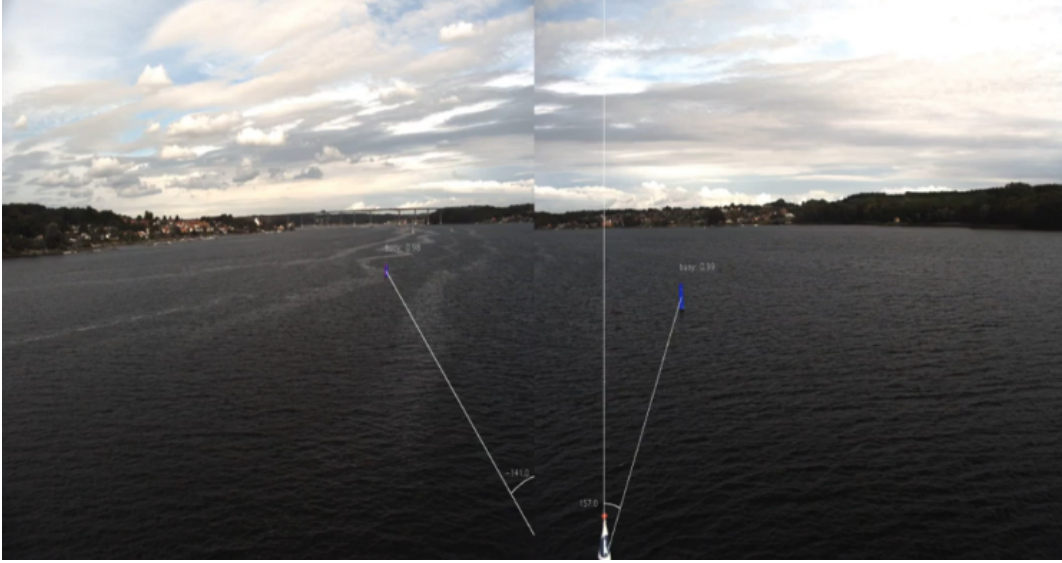


Fig. 9: Object detection and classification on two RGB images are shown by highlighting the detected object in green colour and showing the bearing to detected objects.

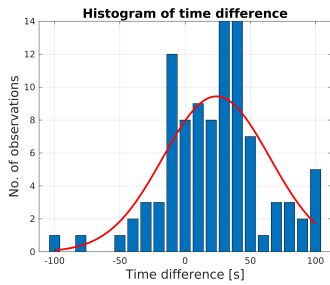


Fig. 10: Histogram of time differences between observations done by the human and electronic lookout (calculated by (1)). The imposed normal distribution has the following parameters: $\mu = 23.9$ s and $\sigma = 41.0$ s. Electronic outlook classifies objects earlier than the human eye fixation by 24s in average.

outlook could serve as a *fifth sense* for the navigator and perhaps pave the way for temporally unmanned bridge in conditions with little other traffic.

10. Conclusions

This study compared human outlook with electronic. Using instance of fixation of eye-tracking glasses with

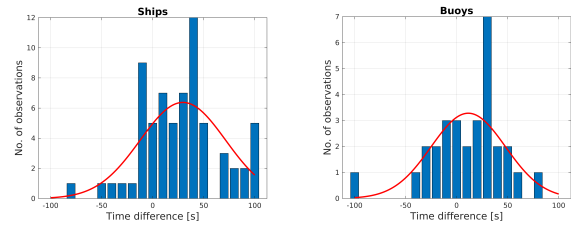


Fig. 11: Histogram of time differences between observations done by the human lookout and the electronic lookout (calculated by (1)). In mean, the electronic outlook detects and classifies objects 30 s faster for ships and 11 s for buoys, compared to human eye fixations. Negative outliers should be avoided by improving robustness.

instance of electronic outlook by cameras and mask-RCNN classification, the study provided statistics for a comparison on one of the essential parameters. The performance of the Mask-RCNN was evaluated on the validation set of annotated RGB images. Object detection performance showed a satisfactory detection probability for objects larger than 400-500 pixels in an image, a quantification that is useful for camera system design for electronic outlook. Some outliers were found to exist in

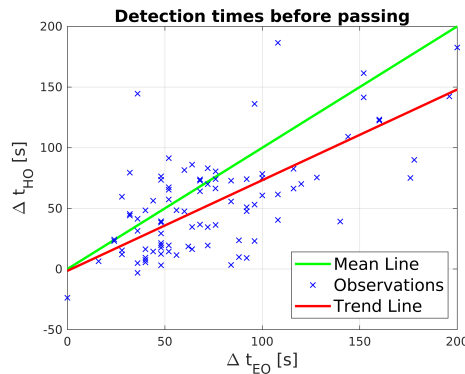


Fig. 12: Scatter diagram of time to react. The plot shows the range 0 – 200 s. The trend line shows that time to react is longer with electronic outlook than time after a fixation.

form of false detections. A single instance of missed detection was also found in the validation data. Robustification of the classifiers will be needed to obtain the required dependability of electronic outlook and is a topic of further research.

Acknowledgments

The authors would like to acknowledge the dedicated efforts made by laboratory engineers, present and former students. The participation of the ferries: MF Isefjord, MS Pernille, MF Højestene and MF Marstal is gratefully acknowledged. This research was initiated by the Danish Maritime Authority and funded by the Danish Maritime Foundation via DTU's Maritime Centre. This funding is gratefully appreciated.

References

- [1] *Tobii Pro Glasses 2, User's Manual*, 1.0.5 edition, 2018.
- [2] S. Aleem. *Analysis of shiphandlers' eye-gaze and simulation data for improvements in COVE-ITS system*. PhD thesis, Naval Postgraduate School, Monterey, California, 2017.
- [3] F. Bjørneseth, C. Loraine, M. Dunlop, and S. Komandur. Towards an understanding of operator focus using eye-tracking in safety-critical maritime settings. In *Proc. Int. Conference on Human Factors in Ship Design & Operation*, Glasgow, February 2014.
- [4] Domenico D Bloisi, Andrea Pennisi, and Luca Iocchi. Background modeling in the maritime domain. *Machine vision and applications*, 25(5):1257–1269, 2014.
- [5] Fouad Boussetouane and Brendan Morris. Fast CNN surveillance pipeline for fine-grained vessel classification and detection in maritime scenarios. In *Advanced Video and Signal Based Surveillance (AVSS), 2016 13th IEEE International Conference on*, pages 242–248. IEEE, 2016.
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.
- [7] Sergiy Fefilatyev, Dmitry Goldgof, Matthew Shreve, and Chad Lembke. Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system. *Ocean Engineering*, 54:1–12, 2012.
- [8] F. Forsman, A. Sjors, J. Dahlman, T. Falkmer, and H. C Lee. Eye tracking during high speed navigation at sea. *Journal of Transportation Technologies*, 277, 2012.
- [9] Ross Girshick. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015.
- [10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
- [11] O. S. Hareide and R. Ostnes. Maritime usability study by analysing eye tracking data. *The Journal of Navigation*, 70(5):927–943, 2017.
- [12] O.S. Hareide and R. Ostnes. Scan pattern for the maritime navigator. *Transnav – the International Journal on Marine Navigation and Safety of Sea Transportation*, 2017.
- [13] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [14] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer. *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford University Press, 2011.
- [15] Maxime Leclerc, Ratnasingham Tharmarasa, Mihai Cristian Florea, Anne-Claire Boury-Brisset, Thiagalingam Kirubarajan, and Nicolas Duclos-Hindé. Ship classification using deep learning techniques for maritime target tracking. In *2018 21st International Conference on Information Fusion (FUSION)*, pages 737–744. IEEE, 2018.
- [16] Frederik S Leira, Tor Arne Johansen, and Thor I Fossen. Automatic detection, classification and tracking of objects in the ocean surface from UAVs using a thermal camera. In *Aerospace Conference, 2015 IEEE*, pages 1–10. IEEE, 2015.
- [17] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [18] Rodrigo Da Silva Moreira, Nelson Francisco Favilla Ebecken, Alexandre Soares Alves, Frédéric Livernet, and Aline Campillo-Navetti. A survey on video detection and tracking of maritime vessels. *International Journal of Recent Research and Applied Studies*, 20(1), 2014.
- [19] B. Muczynski, M. Gucma, M. Bilewski, and P. Zalewski. Using eye tracking data for evaluation and improvement of training process on ship's navigational bridge simulator. *Maritime University of Szczecin*, 33(105):75–78, 2013.
- [20] M. Pico, Hoogerveg D., R. Bik, S. van der Wiel, and R. van Basten Batenburg. Enhancing situational awareness. a research about improvement of situational awareness on board of vessels. Technical report, . Rotterdam Mainport University of Applied Sciences, 2015.
- [21] Dilip K Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey. *IEEE Transactions on Intelligent Transportation Systems*, 18(8):1993–2016, 2017.
- [22] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, Jan 2017.
- [23] S.K. Renganayagalu and S. Komandur. Video support tools for training in maritime simulators. In *Proc. of the International Conference on Contemporary Ergonomics and Human Factors*, Cambridge UK, April 2013.
- [24] Bryan C. Russell, Antonio Torralba, Kevin P. Murphy, and William T. Freeman. Labelme: A database and web-based tool for image annotation. *Int. J. Comput. Vision*, 77(1-3):157–173, May 2008.
- [25] M. Stadler. *Psychology of Sailing. The sea's effects on mind and body*. Adlard Coles Ltd, 1984.
- [26] Zygmunt L Szpak and Jules R Tapamo. Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set. *Expert systems with applications*, 38(6):6669–6680, 2011.
- [27] F. C. van Westrenen. *The Maritime Pilot at Work. Evaluation and use of a time-to-boundary model of mental workload in human-machine systems*. Dr.ing. thesis, The Netherlands TRAIL Research School, 1999. TRAIL Thesis Series nr T99/2.
- [28] Yang Zhang, Qing-Zhong Li, and Feng-Ni Zang. Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean Engineering*, 141:53–63, 2017.